Rec'd PCT/PTO 13 JAN 2005

10/521495

FUNCTIONAL SCREENING METHOD

Technical Field

5   The present invention relates to novel high-throughput functional genomic methods for determining gene and protein function in a cellular context. The method also has utility in identifying novel chemical modulators of gene and protein /enzyme activity.

10   **Background to the Invention**

The large amounts of gene sequence, gene expression and protein expression data arising from the Human Genome Project, and from further downstream investigative efforts, have the potential to allow identification of

15   many new drug targets. Realisation of this potential will require significant efforts in determining the function of new gene products and validating these proteins as drug targets.

Obtaining valid functional information on gene and protein function

20   requires function to be determined (or confirmed) in-context; i.e. the function of the gene/protein should be determined in the presence of other genes/proteins which are likely to interact with it. Consequently there is a need for cell-based approaches for functional screening that enable functional information to be derived in-situ in a cellular environment where dynamic

25   interactions between components may require other cellular components not available in a solution assay.

Moving high-throughput biology into cellular assays can build on and parallel previous work correlating and clustering transcription and interaction

30   data derived from micro-array and protein-protein interaction studies (Ge et al. (2001) Nature Genetics 29, 482-486). Aided by high throughput analysis technologies, cellular screening based approaches can begin to address the

1

complexity necessary to unravel intracellular pathways and control in
mammalian cells (Giese *et al.* (2002) Drug Discovery Today 7, 179-185), with
the ultimate aim of producing sufficiently detailed description to allow a
representation of cellular processes at a system level (Endy & Brent (2001)
5   Nature *409*, 391-395; Kitano (2002) Science *295*, 1662-1664).

To achieve functional screening in a cellular context two elements are
required;

10   a) genetic effector(s) or chemical modulator(s)
      b) measurable phenotype(s); i.e. an assay read-out from a test system.

to establish a cause-and-effect relationship between genes and phenotype or
between chemicals and phenotypes. These elements can be used in a variety
15   of screening processes that differ only in their objectives:

      1) functional genomics; discovery of gene function in normal biology
      2) target validation; discovery of gene function in aberrant biology
      3) chemical genetics; discovery of chemicals that modulate normal
20   phenotypes
      4) drug discovery; discovery of chemicals that modulate aberrant phenotypes

In current procedures a test system is interrogated for the effects of
genetic or chemical variance (i.e. up- or down-regulating expression of one
25   gene, or the presence or absence of a candidate drug respectively), either
alone or in combination. Consequently the effects (and by inference the
function) of a gene (*effector*) or a drug (*modulator*) on a read-out from a test
cell can be measured in isolation or in combination by observation of the
behaviour of the test system. By using combinations of effectors and
30   modulators of known and unknown function it is possible to begin to derive
functional linkage between known and unknown entities and hence to assign
function.

Use of cell-based assays for such screens (Croston (2002) Trends in Biotechnology *20*, 110-5; Zheng & Chan 2002 Current Issues in Molecular Biology *4*, 33-43) is becoming more widely adopted for reasons of acquiring

5   contextual information as described above. Such assays employ a wide variety of assay methodologies, including reporter gene assays, cell growth, pre-cursor incorporation, cell transformation, cell morphology, and fluorescent enzyme assays. These approaches to functional screening have typically used pre-existing assays and instrumentation (e.g. measurement of a

10   luciferase reporter gene in a luminometer) which require assay development prior to the screening process and which yield data averaged for a cell population under test.

US 6322973 (Iconix Pharmaceuticals) describes surrogate means for

15   discovering chemical modulators of genes of unknown function. A heterologous gene of unknown function is expressed in a host cell (e.g. expression of a human gene in a yeast cell) and the host cell is evaluated for a resulting change in phenotype which can then be used as the basis of a cellular assay. Consequent exposure of the host cell exhibiting an altered

20   phenotype to a test substance and assaying for an effect of the test substance on the cellular assay identifies test substances which are modulators of the function of the heterologous gene.

US 6340595 (Galapagos Genomics) describes means for identifying the

25   function of the products of a library of sample nucleic acids by expression of the library of nucleic acids in adenoviral vectors. The sample nucleic acids are synthetic oligonucleotides, DNA, or cDNA and encode polypeptides, antisense nucleic acids, or genetic suppressor elements. The sample nucleic acids are expressed in a host and the resultant altered phenotype used to assign a

30   biological function to the product encoded by the sample nucleic acid.

3

WO0202740 (Rosetta Inpharmatics) describes methods and systems (e.g., computer systems and computer program products) for characterising cellular constituents, particularly genes and gene products. The invention provides methods for assigning or determining the biological function of

5   uncharacterised genes and gene products by using response profiles derived from measurements of pluralities of cellular constituents in cells having a modified gene or gene product, as phenotypic markers for the gene product. Methods are provided for clustering such response profiles so that similar or correlated response profiles are organised into the same cluster. The

10   invention also provides databases of response profiles to which the response profile of an uncharacterised gene or gene product are compared.

       WO0171023 (Genetrace) describes methods for deciphering genetic function. The method provides a matrix of cell lines in which target-specific

15   modified cell lines differ from parental cells in the activity or concentration of a selected protein or nucleic acid. The matrix of cells is exposed to one or more stimuli or test compounds and the cell matrix profiled for response(s) to the stimuli or test compounds. Analysis of the resulting profiles yields information on the genetic function of elements that differ in activity or concentration

20   across the matrix of cells.

       All of the above prior-art methods are characterised by one or more of the following;

       a) measurement of the effects of heterologous genes (e.g. human genes in

25   yeast)

       b) a requirement for development of suitable assays prior to screening

       c) a requirement for engineered cell lines prior to screening.

       A significant problem encountered in the prior art assays described above

30   is that they rely on pre-existing assays and are thus, *a priori*, limited in scope, coverage of biological events being limited by the availability of known assays. This leads to the further problem that assignment of function is limited to those

4

entities which interact with a biological process linked to an available assay read- out. Furthermore, since in general these assays report on cause and effect relationships averaged across a cell population, they do not yield information on the distribution of response across a cell population (e.g. due

5   to cell cycle status, or due to a mixed population of responding and non-responding cells).   An additional problem with the prior art methods is that the assays can only be used on stable populations of cells and are not generally suitable for use with non-homogeneous populations of cells such as transiently transfected cells.

10

Consequently what is required to increase the efficiency of functional screening are methods which do not require pre-existing assays, have the broadest possible coverage of cellular processes and provide data at the individual cell level. The present invention provides methods for functional

15   screening in which assays are generated in concert with screening in an iterative process which expands the scope of biological coverage with each iteration and which uses image-based analysis to yield data at sub-cellular resolution.

20   The method of the present invention circumvents at least some of the limitations of prior-art methods discussed above by providing means to generate functionally diagnostic assays which are integrated into a functional screening process. The method takes advantage of the fact that many cellular proteins exhibit a characteristic cellular localisation and in many cases change

25   their cellular localisation in response to certain stimuli. Consequently, given collections of coding nucleic acid sequences and of chemical compounds, where both collections contain members of known and unknown function, it is possible to generate pairings of one nucleic acid sequence with one chemical compound to produce a specific cellular localisation of a marker coupled to

30   the product of the nucleic acid sequence. Such pairings may then be used as diagnostic assays for testing against other collection members and thus build up clusters and linkages therebetween. In this way, using some members of

each collection which are of known function, it is possible to assign function to previously uncharacterised elements by linkage to known elements.

Thus the method of the present invention allows function to be assigned at a
5   molecular and temporal level for any cellular component, chemical, drug or other active moiety which induces a change in behaviour of an endogenous or exogenous cellular component by reference to changes induced by other moieties of known function. Non-destructive single cell analytical methods are used to analyse the cellular behaviour of indicators influenced by genetic effectors and chemical modulators, where the indicators
10 and effectors may be either endogenous or exogenous to the cell.

Summary of the Invention

According to a first aspect of the present invention, there is provided a method for
15 determining the function or effect of a genetic element or a chemical modulator from a library of genetic elements and chemical modulators of known and unknown function on a population of cells, the method comprising

i)   determining the distribution of an indicator nucleic acid sequence being expressed in the cells in the presence and the absence of a first chemical
20   modulator, which modulator affects the distribution of the indicator, wherein the cells are both co-expressing an effector nucleic acid sequence and are in the presence of a second chemical modulator; and

ii)   analysing the distribution data from all combinations of the effector,
25   modulator and indicator to derive functional linkages and assign function to the effector and the second modulator.

In the context of the present invention, the following terms are to be interpreted as defined below:

6

'*Effector*' – a nucleic acid sequence with biological function or activity, resulting either from an expressed protein with biological function or activity (e.g. cDNA or other coding nucleic acid sequence) or resulting from another mechanism of action (e.g. antisense and RNAi sequences);

5   '*Modulator*' – a chemical moiety with biological function or activity;

'*Indicator*' - a nucleic acid sequence which comprises a detectable label, encodes a detectable label or which may optionally be fused to a sequence encoding a detectable protein label and expressed in a cell resulting in a characteristic localisation of the detectable protein;

10   '*Cellular Assay*' - an assay providing a diagnostic read-out of the biological activity of an effector or modulator

In a second aspect of the present invention, there is provided a method for determining the function or effect of a genetic element or a chemical modulator from a
15   library of said genetic elements and chemical modulators of known and unknown function on a population of cells, the method comprising

i)   determining the distribution of an indicator nucleic acid sequence being expressed in said cells in the presence of a first chemical modulator, which
20   modulator affects the distribution of the indicator, wherein the cells are both co-expressing an effector nucleic acid sequence and are in the presence of a second chemical modulator;

ii)   comparing the distribution data of i) above with known distribution data,
25   stored on an electronic or optical database, for the indicator nucleic acid sequence in the absence of the first chemical modulator; and

iii) analysing the distribution data from all combinations of the effector, modulator and indicator to derive functional linkages and assign function to the
30   effector and the second modulator.

7

Suitably, the effector nucleic acid sequence encodes a protein or peptide and is selected from the group consisting of DNA, cDNA, RNA and Protein Nucleic Acid.

Preferably, the effector nucleic acid sequence is an antisense oligonucleotide (cf.
5   Dean (2001) Current Opinion in Biotechnology, *12*, 622-625). More preferably, the effector nucleic acid is a small interfering RNA (siRNA) which causes gene silencing (cf. Elbashir *et al.* (2002) Methods, *26*, 199-213). RNA interference (RNAi) is a highly conserved gene silencing mechanism that uses double-stranded RNA as a signal to trigger the degradation of homologous mRNA. The mediators of sequence-specific
10   mRNA degradation are 21- to 23-nt small siRNAs generated by ribonuclease III cleavage from longer double-stranded RNA.

Preferably, there is provided an expression vector comprising suitable expression control sequences operably linked to an indicator or an effector nucleic acid sequence .
15   according to the present invention. The DNA construct of the invention may be inserted into a recombinant vector, which may be any vector that may conveniently be subjected to recombinant DNA procedures. The choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, ie. a vector which exists as an extrachromosomal entity, the replication of which
20   is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

8

AMENDED SHEET

The vector is preferably an expression vector in which the effector or indicator nucleic acid sequence is operably linked to additional segments required for transcription of the nucleic acid. In general, the expression vector is derived from plasmid or viral DNA, or may contain elements of both.

5   Preferably, the expression vector is selected from the group consisting of plasmid, retrovirus and adenovirus. The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through to protein synthesis.

10

The promoter may be any DNA sequence which shows transcriptional activity in a suitable host cell of choice, (eg. a mammalian cell, a yeast cell, or an insect cell) for transcription of the indicator or effector nucleic acid sequence. The promoter may be derived from genes encoding proteins either

15   homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the nucleic acid sequences of the invention in mammalian cells are the CMV promoter (US 5168062, US5385839), Ubiquitin C promoter (Wulff *et al.*(1990)

20   FEBS Lett. *261*, 101-105), SV40 promoter (Subramani *et al.*(1981) Mol. Cell Biol. *1*, 854-864) and MT-1 (metallothionein gene) promoter (Palmiter *et al.* (1983) Science *222*, 809-814). An example of a suitable promoter for use in insect cells is the polyhedrin promoter (US 4745051; Vasuvedan *et al.*(1992) FEBS Lett. *311*, 7-11). Examples of suitable promoters for use in yeast host

25   cells include promoters from yeast glycolytic genes (Hitzeman *et al.*(1980) J. Biol. Chem. *255*, 12073-12080; Alber & Kawasaki (1982) J. Mol. Appl. Gen.*1*, 419-434) or alcohol dehydrogenase genes (Young *et al.*, in Genetic Engineering of Microorganisms for Chemicals (Hollaender *et al*, eds.), Plenum Press, New York, 1982), or the TPI1 (US 4599311) or ADH2-4c (Russell *et*

30   *al.*(1983) Nature *304*, 652-654) promoters.

The effector and indicator nucleic acid sequences of the present invention may also, if necessary, be operably connected to a suitable terminator, such as the human growth hormone terminator, TPI1 or ADH3 terminators. The vector may further comprise elements such as

5   polyadenylation signals (e.g. from SV40 or the adenovirus 5 Elb region), transcriptional enhancer sequences (e.g. the SV40 enhancer) and translational enhancer sequences (e.g. the ones encoding adenovirus VA RNAs).

10   The vector may further comprise a DNA sequence enabling internal ribosomal entry and expression of two proteins from one bicistronic transcript mRNA molecule. For example, the internal ribosomal entry sequence from the encephalomyocarditis virus (Rees S, et al. (1996) BioTechniques, 20, 102-110 and US 4937190).

15   The recombinant vector may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. An example of such a sequence (when the host cell is a mammalian cell) is the SV40 origin of replication.

20   When the host cell is a yeast cell, examples of suitable sequences enabling the vector to replicate are the yeast plasmid 2µ replication genes REP 1-3 and origin of replication.

25   The vector may also comprise selectable markers, such as a gene that confers resistance to a drug, e.g. ampicillin, kanamycin, tetracyclin, chloramphenicol, puromycin, neomycin or hygromycin.

The procedures used to ligate the effector and indicator nucleic acid

30   sequences of the invention, the promoter and optionally the terminator and/ or targeting sequence, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to

persons skilled in the art (e.g. Molecular Cloning, Sambrook & Russell, Cold Spring Harbour Press 2001).

5   Suitably, the indicator nucleic acid sequence comprises a detectable label or encodes a detectable label. Preferably, indicator nucleic acid sequence is created by fusing the effector sequence to a nucleic acid sequence encoding a detectable label.

10   Suitably, the detectable label is selected from the group consisting of fluorescent protein, enzyme, antigen and antibody.

Fluorescent proteins and fluorescent protein derivatives of chromoproteins have been isolated from a wide variety of organisms, including *Aequoria victoria, Anemonia* species such as *A. majano* and *A.*
15   *sulcata, Renilla* species, *Ptilosarcus* species, *Discosoma* species, *Claularia* species, *Dendronephthyla species, Ricordia* species, *Scolymia* species, *Zoanthus* species, *Montastraea* species, *Heteractis* species, *Conylactis* species and *Goniopara* species.

20   The use of Green Fluorescent Protein (GFP) derived from Aequorea victoria has revolutionised research into many cellular and molecular-biological processes. However, as the fluorescence characteristics of wild type (native) GFP (wtGFP) are not ideally suited for use as a cellular reporter, significant effort has been expended to produce variant mutated forms of GFP
25   with properties more suitable for use as an intracellular reporter (Heim *et al.*, (1994), Procedings of the National Acadamy of.Sciences (USA), *91*, 12501;. Ehrig *et al.*, 1995, FEBS Letters, *367*,163-6; WO96/27675; Crameri, A. *et al.*, (1996), Nature Biotechnology *14*, 315-9; US 6172188; Cormack, B.P. *et al.*, (1996) Gene *173*, 33-38; US 6194548; US 6077707 and GB Patent Number
30   2374868 ('Amersham Biosciences UK Ltd.'). Preferred embodiments disclosed in GB Patent No 2374868 comprise GFP derivatives selected from

the group consisting of: F64L-V163A-E222G-GFP, F64L-S175G-E222G-GFP, F64L-S65T-S175G-GFP and F64L-S65T-V163A-GFP.

5   In a preferred embodiment, the fluorescent protein is a modified Green Fluorescent Protein (GFP) having one or more mutations selected from the group consisting of Y66H, Y66W, Y66F, S65T, S65A, V68L, Q69K, Q69M, S72A, T203I, E222G, V163A, I167T, S175G, F99S, M153T, V163A, F64L, Y145F, N149K, T203Y, T203Y, T203H, S202F and L236R.

10   Preferably, the modified GFP has three mutations selected from the group consisting of F64L-V163A-E222G, F64L-S175G-E222G, F64L-S65T-S175G and F64L-S65T-V163 as disclosed in GB Patent Number 2374868.

15   Preferably, the enzyme is selected from the group consisting of β-galactosidase, nitroreductase, alkaline phosphatase and β-lactamase. The indicator nucleic acid sequence can thus be detected by the action of the enzyme on a suitable substrate added to the cell. Examples of such substrates include nitro-quenched CyDyes™ (Amersham Biosciences, nitroreductase substrate), ELF 97 (Molecular Probes, alkaline phosphate 20 substrate) and CCF2 (Aurora Biosciences, β-lactamase substrate).

Suitably, the modulator is selected from the group consisting of organic compound, inorganic compound, peptide, polypeptide, protein, carbohydrate, lipid, nucleic acid, polynucleotide and protein nucleic acid. Preferably, the 25 modulator is selected from a combinatorial library comprising similar organic compounds such as analogues or derivatives.

Suitably, the cell is a eukaryotic cell. Preferably, the eukaryotic cell is selected from the group consisting of mammal, plant, bird, fungus, fish, insect 30 and nematode, which cell may or may not be genetically modified. More preferably, the mammalian cell is a human cell, which cell may or may not be genetically modified.

Preferably, the localisation of the detectable label is determined using an imaging system. A suitable Imaging System is the In Cell Analyzer, as described in WO 99/47963 and PCT/GB03/01816.

5   In a third aspect of the present invention, there is provided an automated system for determining the function or effect of a chemical and /or a genetic element on a population of cells comprising use of the method as hereinbefore described together with an imaging system and a computerised data processing device.

10

Brief Description of the Invention

Figure 1; Schematic for generation of an indicator cell assay from a cDNA collection.

15   Figure 2; Schematic for establishing an inferred functional relationship between an effector and a modulator in a cellular assay.

13

Figure 3; Schematic for generation of an indicator assay from a cDNA collection and a chemical collection and subsequent application of selected indicator assays for establishing functional relationships between components of the two collections.

5

Figure 4: a) Triplet functional relationship between effector, modulator and indicator. b) variation in triplets derived from effector and modulator collections comprising components of known and unknown function and/or biological activity.

10

Figure 5; Schematic for establishing extended functional relationships between effector and/or modulators of known and unknown function through connection of triplet functional relationships through common components.

15   Figure 6; Image fluorescence intensity measurements for a nuclear DNA stain and EGFP-fusion protein expression for a range of cDNA indicators transfected into HeLa cells.

Figure 7; Image fluorescence intensity measurements for a nuclear DNA stain
20   and EGFP-fusion protein expression from a single cDNA indicator transfected into HeLa cells.

Figure 8; Nuclear:cytoplasmic indicator distribution in HeLa cells exposed to dexamethasone and staurosporine.

25

Figure 9; Scatterplot of indicator distribution in HeLa cells exposed to dexamethasone and staurosporine.

Figure 10; Response of a range of indicators to staurosporine exposure of
30   HeLa cells.

Figure 11; Effects of transient transfection of a range of cDNA effectors on distribution of a NFκB p65-GFP indicator in CHO cells.

Figure 12; Effects of transient transfection of a range of cDNA effectors on
5   the response of a NFκB p65-GFP indicator to IL-1 stimulation in CHO cells.

Figure 13; Effects of transient transfection of a range of cDNA effectors on distribution of a Rac1-GFP indicator in CHO cells.

10

Detailed Description of the Invention

To achieve the method of the current invention one or more of a
15   collection of nucleic acid sequences [10] (Figure 1) in a vector suitable for
expression of the nucleic acid in a host cell are subcloned into a further vector
[20] to provide fusions of the protein product of the nucleic acid sequence(s)
with a detectable protein. The detectable protein may be any protein which
may be expressed in a mammalian cell and detected using appropriate
20   instrumentation. Suitable detectable proteins include fluorescent proteins such
as Green Fluorescent Protein   Expression of the fusion protein in mammalian
cells may be achieved by use of standard methods including chemically
mediated transfection (FuGENE, Roche; Lipofectin, Invitrogen),
electroporation (Brunner et al. (2002) Molecular Therapy 5, 80-6) or ballistic
25   delivery (Burkholder et al. (1993) J Immunol Methods 165,149-56).

Expression of the detectable fusion protein in a population of host cells
[30] yields a distribution of the detectable protein characteristic of the
distribution of the protein encoded by the nucleic acid sequence [10].
30   Expression of the fusion protein in a second population of host cells [50] in the
presence of a test compound [40] will in certain circumstances yield a
distribution of the fusion protein [70] which differs from that in the absence of
the test compound [60]. In such cases of combinations of [20] and [40] which

yield distribution patterns where [60] differs from [70] the particular
combination of test compound and detectable fusion protein provide a basis
for further investigations. It is an important aspect of this process that it does
not require knowledge of the identity or biological function of either component

5   [10] or [40] to be known, beyond that required to follow the process as
described; e.g. sufficient sequence information for [10] to enable assembly of
the fusion construct [20]. This process establishes combinations of fusion
proteins [20] and test compounds [40] which together engineer a defined and
responsive cellular phenotype, i.e. a cell-based assay which can be used in

10   further functional screening.

Once key combinations of [20] and [40] have been established in which
[40] exhibits a reproducible activity in modulating the cellular distribution of
[20], a second round of screening may be carried out in which nucleic acid

15   sequences [10] are transfected into cells expressing the detectable fusion
protein in the absence [60] and presence [70] of the test compound [40]. Cells
are subsequently evaluated for modulation of the engineered phenotype to
identify nucleic acid sequences [10] which modulate the cellular distribution of
the detectable fusion protein either alone [80], or in combination [90]

20   (antagonism or synergy) with the test compound.

Repetition of the screening process (Figure 2) using libraries of nucleic
acid sequences [110] and test compounds [140], where both libraries contain
elements of known (shaded) [111] [141] and unknown (unshaded) function

25   [112] [142], and exposing cells of engineered phenotype to elements of these
libraries alone [160] [162] and in combination [165], allows the functions and
interactions of nucleic acid sequences and test compounds to be investigated.
In the example of Figure 2, interaction of a nucleic acid sequence component
[170, 166, 168] of the library [110] with cells of engineered phenotype [160]

30   causes a change in the detected phenotype [170]; interaction of a chemical
component of the test compound collection [140] with cells of the same
engineered phenotype [162] does not change the detected phenotype [166];

co-exposure of further cells of the same engineered phenotype [165] to the same chemical and genetic elements in combination does not lead to a change in the observed phenotype [168], indicating some form of antagonism between the functions of the test compound and the expressed nucleic acid
5   sequence.

Large scale screening using a library of nucleic acid sequences of known and unknown function in combination with a collection of test compounds of characterised or uncharacterised biological activity can
10   therefore be carried out to establish combinations of nucleic acid sequences and chemical entities which operate in concert to modify a detectable cellular phenotype measured by a cellular assay. Since the process inherently generates cellular assays the method has advantages over previously used approaches in that it does not require either prior knowledge of biological
15   activities or pre-existing cell assays; although the process may be used in conjunction with pre-existing cell-based assays, where available.

A number of groups (Bejarano *et al.* (1999) J Cell Sci *112 (23)*, 4207-11; Misawa *et al.(* 2000) Proc Natl Acad Sci U S A *97,* 3062-6; Gonzalez *et*
20   *al.*(2000) Trends Cell Biol *10,*162-5; Rolls *et al* .(1999) J. Cell Biol. *146,* 29-44; Simpson *et al.* (2000) EMBO *1,* 287-92) have reported using GFP tagging of unknown genes or sequence motifs arising from cDNA libraries or other sources to identify sequences associated with proteins of defined sub-cellular localisation. Developments are already in place to automate cloning (Rolls *et*
25   *al.* (1999) J. Cell Biol. *146,* 29-44) which allows high-throughput generation of the N- and C-terminal GFP fusions necessary for transfection.

Use of high throughput image based analysis using instruments such as the Amersham Biosciences IN Cell Analyzer (Goodyer *et al.* (2001),
30   Society for Biomolecular Screening, 7th Annual Conference and Exhibition, Baltimore, USA *Screening and signalling events in live cells using novel GFP redistribution assays*) permits the use of assays measuring tagged protein

localization to be carried out on transiently modulated cells, (e.g. by transient cDNA transfection) with data collected on an individual cell basis. This approach offers a number of benefits, including removal of the need to pre-establish stable indicator cell lines prior to screening yields assay results

5   which are less likely to be distorted by 'over-expression squelching' and phenotype distortion arising through cellular selection (Giese *et al* Drug Discovery Today (2002) 7, 179-186) associated with generation of large numbers of stable cell lines.

10   The method of the invention may be used to establish functional relationships between genetic elements (effectors), chemical elements (modulators) and cellular assays (indicators). Starting from collections of effectors [210] (Figure 3) and modulators [240] of known or unknown function, cDNA effectors are engineered as fusions with a detectable marker protein

15   [220] and transfected into target cells in the presence [270] and absence [260] of selected modulators [240]. Combinations of effectors, modulators and target cells giving a reproducible difference in the localisation of the detectable fusion protein are selected [S] for further rounds of functional screening in which the selected combinations are challenged with effectors [210] or

20   modulators [240]. By this means many three-way combinations of effectors, modulators and indicators may be tested [290]. Tri-partite combinations [390] (Figure 4a) in which the activity [345] of a chemical modulator [340] and the activity [315] of a genetic effector [310] on a indicator cell based assay [360] are correlated and used to infer the presence or absence of a functional

25   linkage [301] between effector and modulator, may be used to establish functional links and clusters between many different entities. For any collections of effectors and modulators where the biological function or activity of components of the collections are both known and unknown, and where these collections are tested in combination with indicator cell assays of a

30   known (i.e. pre-existing assays) or unknown biological significance, eight possible three-way combinations (triplets) are possible [302]-[309], and are summarised in Table 1.

Consequently by collecting data from a large number of triplets where unknown elements are tested in combination with known elements and selecting triplets in which there is an interaction between all three components

5   it is possible to assemble networks of functional linkages which yield information on the biological function of previously uncharacterised elements. For example a triplet [400] (Figure 5), in which the biological activities of both effector and modulator elements are unknown, can be linked to a second triplet [401], in which the biological activity of both modulator and effector are

10   known, through a common assay shared by both triplets, and consequently yields information on the possible biological activities of the modulator and effector of the first triplet [400]. By extension of the same principle triplet [402] can be linked to triplet [401] through a common modulator and further linkages to triplets [403] through [408] established. In Figure 5 such linkages are

15   represented in a two dimensional plane, in practice linkages are not constrained to a linear branching structure and may comprise loops [L1] making further connections, branch point (B) or multiple branch points (e.g. B1, B2) from the same triplet.

20   **Specific Examples**

Example 1

A collection of cDNAs (Invitrogen & Image Consortium, Table 2) were

25   prepared for expression as cDNA-EGFP fusion proteins by inserting cDNA sequences into the multiple cloning site of pCORON1000-EGFP-N2 and pCORON1000-EGFP-C1 expression vectors (Amersham Biosciences) using standard molecular cloning techniques (Molecular Cloning, Sambrook & Russell, Cold Spring Harbour Press 2001). These vectors direct the

30   expression of fusion proteins comprising the protein encoded by the inserted cDNA sequence fused at their amino and carboxy termini to EGFP in mammalian cells under the control of a constitutively active CMV promoter.

Expression vectors encoding cDNA-EGFP indicators were transiently transfected into HeLa cells growing in wells of 96 well microtitre plates by chemically mediated transfection (Fugene, Roche) and cells incubated under

5   standard growth conditions for 24 hours to permit synthesis of indicator fusion proteins. Cells were subsequently stained with DRAQ 5, a cell permeable nuclear DNA binding dye (Biostatus), to fluorescently mark cell nuclei, and all wells imaged with dual laser excitation (EGFP 488nm, DRAQ 5 633nm) using an IN Cell Analyzer (Amersham Biosciences). Data for green (EGFP) and red

10   (DRAQ 5) fluorescence were collected for all cells (Figure 6) and used to   ,
determine thresholds for data separation of transfected cells (EGFP fluorescence above threshold) from non-transfected cells (EGFP fluorescence below threshold). Representative data from a single cDNA-EGFP fusion protein are shown in Figure 7. A fusion protein derived from full length cDNA

15   encoding the glucocorticoid receptor inserted in pCORON1000-EGFP-N2 was expressed in HeLa cells and analysed as described above. For this indicator protein a threshold of 25 (horizontal dotted line on Figure 7) was used to discriminate data from transfected (>25) and non-transfected cells (<25).

20   Data collection and analysis as described above allows cDNA-EGFP fusion proteins to be used as indicators in transiently transfected cell populations by using data thresholding to distinguish transfected from non-transfected cells, so avoiding the need to engineer stable cell lines required for analysis methods which use population average measurements.

25

Example 2

Indicator proteins derived from a range of cDNAs as described for Example 1 were transfected into HeLa cells and allowed to express for 24

30   hours. Following expression, cells were transferred into serum-free media for 2 hours to allow effects of stimuli from serum factors such as cortisol to decay. Cells were stained with DRAQ 5, imaged as described in Example 1, returned

to complete media and then exposed to 1μM dexamethasone (a synthetic
glucocorticoid agonist) or 1μM staurosporine (kinase inhibitor and apoptosis
inducer) for 5 minutes followed by repeat imaging. Image data were analysed
using a nuclear trafficking algorithm (Amersham Biosciences; (cf. Adie *et al.*

5   (2001) 'The pharmacological characterisation of a GPCR using pH sensitive
cyamine dyes on the LEADseeker Cell Analysis System' Poster, Society for
Biomolecular Screening Conference 10-13th September 2001, Baltimore USA;
Goodyer *et al.* (2001) 'Screening of signalling events in live cells using novel
GFP redistribution assays' Poster, Society for Biomolecular Screening

10   Conference 10-13th September 2001). The alogorithm returns a numerical
description of fluorescence distribution in nucleus and cytoplasm as a ratio
(nuclear fluorescence divided by cytoplasmic fluorescence ; N/C). This
algorithm allows the spatial distribution of cDNA-EGFP fusion proteins to be
quantitated in expressing cells: a low N/C ratio indicating a cytoplasmic

15   location for the indicator protein, a high N/C ration indicating a nuclear
location. Consequently a change in N/C ratio for an indicator protein induced
by a chemical modulator indicates a translocation of the indicator in response
to the modulator. This form of analysis permits screening of combinations of
indicators/chemical modulators for pairings in which the indicator exhibits

20   translocation in response to the modulator, and may serve as the basis for
testing the action of effectors or further modulators on the characterised
response.

Results from this analysis are shown in Figure 8 with differences in N/C

25   ratios in the absence and presence of dexamethasone and staurosporine
plotted for a range of indicator fusion proteins. The results show a diversity of
response across the indicator proteins to the two modulators used in this
example. A indicator protein (GR) constructed by fusion of glucocorticoid
receptor to EGFP showed a very large increase in N/C ratio indicative of a

30   change in localisation of the indicator protein from cytoplasm to nucleus. This
change in localisation is consistent with the well characterised translocation
response of glucocorticoid receptor on exposure to glucocorticoid agonists,

including dexamethasone (Htun *et al.* (1996) Proc Natl Acad Sci USA *93(10),* 4845-50). A number of other indicator proteins showed a significant change in N/C ratio when exposed to either dexamethasone or staurosporine (e.g.ATF1, YKT6)

5

Data from this example are also shown in Figure 9 as a scatterplot of dexamethasone response against staurosporine response. Plotting data in this form highlights differential responses of indicators to modulators; most indicators either do not show a response to either modulator or show an
10   equivalent response to both modulator treatments. When plotted in this manner the data clearly show that two indicators, GR (glucocorticoid receptor) and ATF1 (activating transcription factor 1) show specific and differential responses to the two modulators. The involvement of ATF1 in cellular response to stress has been described previously (Wiggin *et al.* (2002) Mol
15   Cell Biol Apr. *,22(8),* 2871-81) indicating that the ATF1-staurosporine pairing would serve as a suitable test system for studying the activity of effectors or modulators on cellular stress response mechanisms. The data shown in Figure 9 also highlight those indicators which responded to both dexamethasone and staurosporine. These responses are a direct result of the
20   serum removal and replacement regime required to measure GR translocation, where a group of indicator proteins, including CREB1, P27-KIP and LMNA show a change in N/C value following the return of cells to serum containing medium.

25   <u>Example 3</u>

A further group of indicator proteins were transfected into HeLa cells and cells imaged before and after exposure to staurosporine as described in Example 2. Images were analysed with a further two IN Cell Analyzer
30   algorithms, Granularity and Membrane Spot (Amersham Biosciences) (cf. Adie *et al.* (2001) 'The pharmacological characterisation of a GPCR using pH sensitive cyamine dyes on the LEADseeker Cell Analysis System' Poster,

Society for Biomolecular Screening Conference 10-13th September 2001,

Baltimore USA; Goodyer *et al.* (2001) 'Screening of signalling events in live

cells using novel GFP redistribution assays' Poster, Society for Biomolecular

Screening Conference 10-13th September 2001). These algorithms return

5   results which quantitate fluorescence in degrees of granularity (i.e. low value

indicates uniform distribution, high value indicates punctate distribution) and in

terms of membrane localisation. Consequently these algorithms are suitable

for examining indicators which no not exhibit cytoplasmic to nuclear

differential localisation and hence are unsuitable for analysis by the algorithm

10   used in the previous example.

Results from analysis with these two algorithms on staurosporine

treated cells are shown in Figure 10. Data returned by the algorithms varied

significantly across the range of indicators, with some proteins yielding a high

15   granularity value and a low membrane spot value, and vice versa.

Examination of the ratios of the outputs from the two algorithms (Figure 10

inset) revealed that the indicator, Cyt-C (EGFP-Cytochrome C), showed the

highest differential return from the two algorithms. Release of Cytochrome-C

from mitochondria and subsequent cellular redistribution is a well

20   characterised early event in the onset of cellular apoptosis (Gao *et al.* (2001) J

Cell Sci., *114*, 2855-62). Consequently, data from this example provide further

evidence that indicator proteins engineered from cDNAs coding for cellular

proteins fused to a detectable marker and transiently expressed in

mammalian cells provide a means of gaining functional information relevant to

25   the protein encoded by the cDNA; such indicator-modulator pairings are

suitable for use in further functional screening.

Example 4

30   A range of cDNA modulators were transiently transfected into CHO

cells expressing a NFκB p65-GFP fusion protein. This indicator undergoes a

well characterised cytoplasmic to nuclear translocation in response to a

number of stimuli, including exposure to Interleukin-1 (IL-1). Cells were incubated for 24 hours post transfection, stained with DRAQ 5, imaged, and then stimulated with IL-1, followed by repeat imaging. N/C ratios were determined for all images using the algorithm described in Example 2, and a

5   scatterplot (Figure 11) prepared from the data.

In an experiment of this design where two factors (stimulus and effector) may change the behaviour of the indicator, a number of possibilities may occur;

10   a) the effector may decrease the indicator N/C ratio prior to stimulus relative
        to a control value (cells in the absence of effector)

     b) the effector may increase the indicator N/C ratio prior to stimulus relative to
        a control value

     c) the effector may decrease the indicator N/C ratio following stimulus relative
15   to a control value

     d) the effector may increase the indicator N/C ratio following stimulus relative
        to a control value

     all of the above may, depending on their combination, result in a modulation of
     the magnitude of change of the indicator N/C ratio induced by IL-1 stimulus.

20   The scatterplot of Figure 1 represents these scenarios graphically by
     separating results into four quadrants;

Quadrant	Indicator behaviour
lower left	$N/C_0 < control : N/C_{IL-1} < control$
lower right	$N/C_0 > control : N/C_{IL-1} < control$
upper right	$N/C_0 > control : N/C_{IL-1} > control$
upper left	$N/C_0 < control : N/C_{IL-1} > control$

25

In addition, the diagonal dotted line on Figure 11 indicates points of equivalent
N/C ratios, consequently the distance from the line (at 90° to the line) of any

value gives a measure of the overall response of the indicator protein to IL-1 stimulation in the presence of a given effector relative to the absence of the effector. It is clear that the effectors used in this experiment are having a range of effects on the distribution of the indicator protein in changing the N/C

5   ratio before and after IL-1 stimulus and in changing the overall response to IL-1 stimulation.

Figure 12 shows a simplified treatment of these results where only data for IL-1 response (i.e. the difference between N/C_0 and N/C_{IL-1}) are shown. These

10 data indicate a range of responses to transfection with effectors ranging from significant antagonism of IL-1 stimulation (CCND3) to strong agonism (e.g. PRKCs A, Z & E and GSK3B). These agonists have previously been shown to modulate the activity of the NFκB signalling pathway (La Porta *et al.* (1998) Anticancer Res. *18*(4A):2591-7; Hoeflich *et al.* (2000) Nature *406 (6791)*,

15 86-90) confirming the validity of using this approach for functional screening of cDNA effectors against indicators expressed in mammalian cells.

Example 5

20   The functional screen of Example 4 was repeated with a second indicator, RAC1 (T)-GFP, in the presence and absence of stimulation with insulin and analysed using the membrane spot algorithm described in Example 3. As in Example 4 it is clear that the effectors used in this experiment are having a range of effects on the distribution of the indicator

25 protein in changing the cellular distribution of the indicator both before and after insulin stimulus and in changing the overall response to insulin stimulation (Figure 13).

Table 1

	Identity or Function		
	modulator	**effector**	**indicator**
[302]	known	known	unknown
[303]	known	unknown	known
[304]	unknown	known	known
[305]	known	unknown	unknown
[306]	unknown	unknown	known
[307]	unknown	known	unknown
[308]	known	known	known
[309]	unknown	unknown	unknown

Table 2

cDNA	Genbank No.	cDNA	Genbank No.	cDNA	Genbank No.
IKBKG	NM_003639	PDK2	L42451	AGPAT2	AF000237
NFKBIA	M69043	VDAC1	BC008482	ICAM2	X15606
PRKCA	X52479	VDAC2	BC012883	CCR6	U45984
PRKCE	X65293	VDAC3	BC002456	NTRK2	X72958
PRKCZ	L14283	CCND2	M90813	HCK	M16591
MAPK13	AF004709	CCND3	M92287	EPHB2	L41939
MAPK14 (p38)	L35253	RPS6KA2	X85106	KIR3DL2	L41270
MAPK8	L26318	ATF 1	X55544	AP1M2	BC003612
PRKACB	M34181	ATF 4	D90209	APBB1	BC010854
SKP2 (p45)	U33761	CDKN1A	L25610	APP	BC004369
PPM1G	Y13936	HDAC 1	D50405	AQP3	BC013566
FGR (src)	M19722	TFDP1	L23959	CLTA	BC009201
GSK3B	L33801	E2F4	S75174	CLTB	BC006457
OSTF1	U63717	SIX1	X91868	GABRA5	BC011403
BHMT	U50929	ATF5	BC005174	GABRB3	BC010641
HSPA1A (hsp70)	M11717	CREB1	BC010636	GJB2	BC002805
PTPN2	M25393	CREB3	BC010158	KCNH2	BC001914
BHLHB2	AB004066	DUSP4	BC002671	KCNJ8	BC000544
BAD	U66879	E2F6	BC008348	KCNQ2	BC000699
MYBPH	U27266	HDAC3	BC000614	P2RX7	BC011913
ACTB	BC002409	HIF1A	BC012527	STATI2	BC010399
AKT1	BC000479	P27-KIP1	BC001971	OPRD1	NM_000911
ARAF1	BC007514	LMNA	BC000511	PTGIR	NM_000960
ARRB1	BC003636	NUP88	BC000335	AGTR2	NM_000686
ARRB2	BC007427	LAMP2	BC002965	CHRM3	NM_000740
BID	BC009197	GNPAT	AJ002190	CHRM1	NM_000738
FADD	BC000334	RPS9	U14971	EGFR	NM_005228
HSP70	BC002453	PRRG1	AF009242	ARF 1	M36340
HSPCB(hsp90)	BC009206	LOC51035	M68864	ARF 3	M74491
MADH1	BC001878	NOT IN UNIGENE	D14825	ARF 4	M36341
MADH4	BC002379	FLJ13052	M37712	YKT6	U95735
MAPK7	BC007992	PLCG2	BC007565	PITPN	D30036
MDM2	BC009893	RIPK2	AF027706	TOM1	NM_005488
MYCBP	BC008686	GYPB	J02982	TRAM	BC000687
NFATC3	BC001050	PROC	X02750	STAT6	BC004973
PSCD2	BC004361	PTEN	BC005821	TRADD	BC004491
STAT3	BC000627				